

# Analysis of Wireless Sensor Network Protocols using Data Mining Techniques

Siddhi Raut<sup>1</sup>, Lalit Dole<sup>2</sup>, Jayant Rajurkar<sup>3</sup>

<sup>1,2</sup>Department of CSE, G.H.Raisoni College of Engineering, Nagpur, Maharashtra, India

<sup>3</sup>Department of Computer Engineering, MIET, Bhandara, Maharashtra, India

**Abstract**— A structure for extracting and mining wireless sensor data. This structure consists of a new formulation for data mining technique using the association rule by distributed extraction method. The new design express the sequential relations between wireless sensors and association rules using data mining technique for estimating the value of missed events and also to improve the performance. The proposed distributed method is designed with wireless sensor network to improve the network life time by minimizing number of messages required for the mining process. The experiment results have shown that our methodology can reduce the number of exchanged messages at least by 40% compared to the existing other methods. The wireless sensor network automatically collects the information or data. The compressed representation structure (CRS) is able to compress the data and sensor network provides easy and fast collection of data. The mining process of the CRS is compared and analyzed.

**Keywords**— Association rule, Apriori and frequent pattern (FP), clustering, CRS, WSN.

## I. INTRODUCTION

Recent advance in miniaturization design have led to the development of small-sized battery-operated sensors that are capable of detecting ambient conditions such as temperature and sound. Sensors are generally equipped with data processing and communication capabilities. The past few years have witnessed increased in the potential use of WSN such as disaster management, combat field reconnaissance, border protection and security surveillance. Sensors in these applications are expected to be remotely deployed in large numbers and to operate autonomously in unattended environments. Since a WSN is composed of nodes with non-replenish able energy resource, elongating the network lifetime is the main concern .A WSN consists of a number of sensor nodes and a sink. Grouping sensor nodes into cluster has been widely pursued by the research community in order to achieve the network scalability objective. In each cluster, a sensor node is elected, termed as the CH. The CH is responsible for not only the general request but also

receiving the sensed data of other sensor nodes in the same cluster and routing (transmitting) these data to the sink. Therefore, the energy consumption of the CH is higher than of other nodes. In order to balance the energy consumption for elongating the lifetime of this WSN, the CH in a cluster is alternate among sensor nodes. Therefore, the CH selection manner will affect the lifetime of this network. The different application scenario context will follow the different definitions of lifetime.

Each sensor node has its sensing request the capabilities of sensing, computation, and wireless communications. The power consumption of wireless communication between two nodes is based on the transmission distance, which has an exponential increment with the distance. Therefore, the routes of the data transmission to the sink will affect the energy consumption. Since the hierarchical architecture provides more flexibility to handle the data routing problem, it is applied extensively to the WSNs.

A WSN consist of large number of sensor nodes equipped with various sensing devices to observe different phenomenon changes in the real world. A sensor node is composed of typically four units- a) sensing unit: sense the desired data from the interested region. b) Memory unit: - store the data until it is sent for future use. c) Computation unit: - computes the aggregated data d) power unit: - provides power supply for entire process. Since sensor nodes are battery powered devices therefore they have limited energy. WSN are usually deployed in inhospitable terrain such as mountainous region where it's very difficult to recharge or replace batteries. Therefore the main aim of any energy efficient routing protocol is to prolong the network lifetime which is possible by minimizing energy consumption of individual nodes. In addition it is also necessary to ensure that the average rate of consumption of energy by each node is also same. This would ensure that the connectivity needed to transmit data from a source node to sink node is also maintained. Since lifetime of a network is defined as time in which a single node losses its energy and get exhausted. More ever, the transceiver is the major unit of energy consumption in sensor node even when sensor

nodes are in idle state. Therefore sensor nodes must be put to sleep (radio off) if they are not required to transmit or receive data. It is assumed that transceiver, processor and sensing unit can be put to sleep independently. It is assumed that when sensor nodes are put to sleep it means that the transceiver and processor are put to sleep. The challenge is to integrate sleep scheduling scheme with routing protocols for WSNs.

## II. LITERATURE SURVEY

### 2.1 Data Mining Process in WSNs

Data mining in sensor networks is the process of extracting application-oriented models and patterns with acceptable accuracy from a continuous, rapid, and possibly non ended flow of data streams from Sensor networks. In this case, whole data cannot be stored and must be processed immediately. Data mining algorithm has to be sufficiently fast to process high-speed arriving data. The conventional data mining algorithms are meant to handle the static data and use the multistep techniques and multi scan mining algorithms for analyzing static data-sets. Therefore, conventional data mining techniques are not suitable for handling the massive quantity, high dimensionality, and distributed nature of the data generated by the WSNs. Table 2.1 shows the summary of difference between traditional data and WSNs data mining process.

It can be observed from Table 2.1 that traditional data mining is centralized, computationally expensive, and focused on disk-resident transactional data. It directly collects data at the central site which is not bounded by computational resources. In comparison with traditional data-sets, the WSNs data flows continuously in systems with varying update rates. Due to huge amount and high storage cost, it is impossible to store the entire WSNs data or to scan through it multiple times. These characteristics of sensor data and the special design issues of sensor networks make traditional data mining techniques challenging. Hence, it is crucial to develop data mining technique that can analyze and process WSNs data in multidimensional, multilevel, single-pass, and online manner.

	Traditional data	WSNs data
Processing architecture	Centralized	Distributed
Data type	Static	Dynamic
Memory usage	Unlimited	Restricted
Processing time	Unlimited	Restricted
Computational power	High	Weak
Energy	No constraints	Limited
Data flow	Stationary	Continuous
Data length	Bounded	Unbounded
Response time	Non-real-time	Real time
Update speed	Low	High
Number of passes	Multipass	Single

Fig.2.1: Difference between Data processing

Challenges according to the following reasons, conventional data mining techniques for handling sensor data in WSNs are challenging.

#### 2.1.1 Resource Constraint

The sensor nodes are resource constraints in terms of power, memory, communication bandwidth, and computational power. The main challenge faced by data mining techniques for WSNs is to satisfy the mining accuracy requirements while maintaining the resource consumption of WSNs to a minimum.

#### 2.1.2 Fast and Huge Data Arrival

The inherent nature of WSNs data is its high speed. In many domains, data arrives faster than we are able to mine. Additionally, spatiotemporal embedding of sensor data plays an important role in WSNs application. This may cause many classical data processing techniques to perform poorly on spatiotemporal sensor data. The challenge for data mining techniques is how to cope with the continuous, rapid, and changing data streams and also how to incorporate user interaction during high-speed data arrival.

#### 2.1.3 Online Mining

In WSNs, environment data is geographically distributed, inputs arrive continuously, and newer data items may change the results based on older data substantially. Most of data mining techniques that analyze data in an offline manner do not meet the requirement of handling distributed stream data. Thus, a challenge for data mining techniques is how to process distributed streaming data online.

#### 2.1.4 Modeling Changes of Mining Results over Time

When the data-generating phenomenon is changing over time, the extracted model at any time should be up-to-date. Due to the continuity of data streams, some

researchers have pointed out that capturing the change of mining results is more important in this area than the mining results. The research issue is how to model this change in the results.

### 2.1.5 Data Transformation

Since sensor nodes are limited in terms of bandwidth, transforming original data over the network is not feasible. Knowledge structure transformation is an important issue. After extracting model and patterns locally from WSNs data, the output is transferred to the base station. The challenge for data mining technique is how to efficiently represent data and discovered patterns over network for transmission.

### 2.1.6 Dynamic Network Topology

Sensor network deployed in potentially harsh, uncertain, heterogenic, and dynamic environments. Moreover, sensor nodes may move among different locations at any point over time. Such dynamicity and heterogeneity increase the complexity of designing an appropriate data mining technique for WSNs. To address these challenges, researchers have modified the conventional data mining techniques and also proposed new data mining algorithms to handle the data generated from sensor networks. In the following section we have provided the taxonomy of these data mining techniques based on the discipline from which they adopt their ideas.

## 2.2 Taxonomy of Data Mining Techniques for WSNs

A classification scheme for existing approaches designed for mining WSNs data is presented. The highest level classification is based upon the general data mining classes used such as frequent pattern mining, sequential pattern mining, clustering, and classification. Most of the frequent pattern mining and sequential pattern mining approaches have adapted the traditional frequent mining techniques such as the Apriori and frequent pattern (FP) growth-based algorithms to find the association among large WSNs data. Cluster-based approaches have adapted the K-mean, hierarchical, and data correlation-based clustering, based upon the distance among the data point, whereas, classification based approaches have adapted the traditional classification techniques such as decision tree, rule-based, nearest neighbor, and support vector machines methods based on type of classification model that they used. These algorithms have very different and distinct roles; therefore, in order to choose the algorithm for WSNs application, one has to decide in term of these top-level classes.

## III. PROBLEM STATEMENT

Specification provides a complete description of all the functions and specifications, it deals with the CH Selection between the different clusters and finding the Head node between them also concept of sleep scheduling play a vital role in this system for reducing energy consumption.

Our contribution in this paper is as follow:

- Design such system which solves the problem of current system.
- The system will minimize the Energy consumption.
- Study how to collect information and study on gathered information.
- Study different existing network protocols.
- Study different types of feasibility like Energy and power.
- Study Network simulator.

## IV. PROPOSED ALGORIHTM

A wireless sensor network SN consists of  $n$  sensor nodes in which nodes are divided into several clusters. The definition of the life-time is defined as follows, in which  $A(N)$  is the available (alive) nodes in a node set  $N$  and  $|A(N)|$  is the number of nodes in  $A(N)$ .

Type 1: The life-time is evaluated when  $|A(SN)| \geq n$ .

Type 2: The life-time is evaluated when  $|A(SN)| \geq 0$ .

Type 3: Otherwise, not in Type 1 and Type 2 cases is defined as Type 3

Each cluster consists of a number of sensor nodes. The sensor nodes at the same cluster consists a coordinator. The coordinators are divided into several disjoint groups. One of the coordinators at the same group will be chosen as the head of this group. The routine of the sensed data aggregation for a sensor node is from this node to its coordinator, from the coordinator to its head, and then from the head to the sink

### Coordinator Selection (Cluster Head Selection)

In each cluster, termed as  $C_i$ ,  $j$ , one of sensor nodes in  $C_i$ ,  $j$  is elected as the coordinator  $H_{i,j}$ . Moreover, one of the coordinators including  $H_{i,1}$ ,  $H_{i,2}$ ... and  $H_{i,m}$  is elected as the head  $H_i$ . A coordinator  $H$  of a cluster  $C$  is responsible for receiving the sensed data of the other sensor nodes in cluster  $C$  and routing to the sink. The coordinator  $H$  is selected from the sensor nodes in the same cluster  $C$ , where the selection is performed round-by-round. This study proposes 2 methods to select the coordinator.

The following variables are used herein for selecting a coordinator.

- $A(C)$  and  $|A(C)|$ :  $A(C)$  represents the current alive sensor nodes in cluster  $C$  and  $|A(C)|$  represents the number of nodes in  $A(C)$ .
- $e(N_i)$ : the remaining energy of a sensor node  $N_i$
- $t(N_i)$ : the times that sensor node  $N_i$  has been selected as a coordinator

Give a cluster  $C$ , the coordinator  $H$  of  $C$  can be selected from one the following methods.

- Random (R) Selection: Select a node  $N_i$  from  $A(C)$  where the value of  $t(N_i)$  is minimal for all nodes in  $A(C)$ .
- Energy (E) Selection: Select a node  $N_i$  from  $A(C)$ , where the value of  $e(N_i)$  is maximal for all nodes in  $A(C)$ .

The R and E selection methods can individually be used to elect the coordinator  $H$  of a cluster  $C$ . Both of the R selection and E selection method are available for the head selection.

### Routing

The methods, to route the sensed data from the sensor nodes to the sink, are divided into aggregation and non-aggregation methods and are described as follows. Given a cluster  $C_i$ ,  $j$ , for each round, the following operations are performed.

- For the aggregation method, each node in  $C_i$ ,  $j$  periodically sends its sensed data to its cluster coordinator  $H_{i,j}$ . Then, the coordinator  $H_{i,j}$  periodically collects the data from each node in the same cluster and the aggregation of the collected data is sent to the head  $H_i$ . For the head  $H_i$ ,  $H_i$  will collect the data sent from all coordinators  $H_{i,1}, H_{i,2}, \dots$ , and  $H_{i,m}$ , and then send the aggregation of the collected data to the sink.
- For the non-aggregation method, when a sensor node senses the desired data, the sensor node immediately sends the data to its cluster coordinator  $H_{i,j}$ . After the coordinator  $H_{i,j}$  receives the sensed data of a sensor, it immediately, without aggregating other data, sends the data to the head  $H_i$  and the head  $H_i$  then send the data to the sink

### Sleep Scheduling

Basically, there are two classes of energy efficient ad hoc and sensor network routing protocols employing a sleep mode in the literature, cluster-based and flat. Both of them achieve energy efficiency by employing different topology management techniques. This section presents a brief review of these two classes of routing to provide a better understanding of the current research issues in this area. In cluster-based routing protocols, all nodes are

organized into clusters with one node selected to be cluster-head for each cluster. This cluster-head receives data packets from its members, aggregates them and transmits to a data sink.

In some cluster-based routing protocols, the cluster-head assigns TDMA slots to its members to schedule the communication and the sleep mode. Low-Energy Adaptive Clustering Hierarchy (LEACH) is designed for proactive sensor networks, in which the nodes periodically switch on their sensors and transmitters, sense the environment and transmit the data. Nodes communicate with their cluster-heads directly and the randomized rotation of the cluster-heads is used to evenly distribute the energy load among the sensors. Threshold sensitive Energy

Efficient sensor Network protocol (TEEN) is designed for reactive networks, where the nodes react immediately to sudden changes in the environment. Nodes sense the environment continuously, but send the data to cluster heads only when some predefined thresholds are reached. Adaptive Periodic Threshold sensitive Energy Efficient sensor Network protocol (APTEEN) protocol combines the features of the above two protocols by modifying TEEN to make it send periodic data. The cluster-based routing protocols can arrange the sleep mode of each node to conserve energy. However, the high complexity and overhead are incurred. That employs probabilistic based sleep modes. At the beginning of a gossip period, each node chooses either to sleep with probability  $p$  or to stay awake with probability  $1 - p$  for the period, so that all the sleep nodes will not be able to transmit or receive any packet during the period. When an active node receives any packet, it must retransmit the same. All sleeping nodes wake up at the end of each period. All the nodes repeat the above process for every period.

### V. CONCLUSION

In this we have analyzed wireless sensor network protocols CH-Selection and HID. The speed of packets transfers through network increases 40% as compared to protocol Increases the energy efficiency of network. Sleep scheduling used for saving the energy between nodes.

In future scope, we will consider different capabilities of sensor nodes, such as the node with GPS equipment has the capability to know the positions of itself and its neighboring nodes. Adaptively adjusting the period of tree reconstruction depending on the input data rate with a view to further increase the network lifetime.

### REFERENCES

- [1] Jennifer Yick, Biswanath Mukherjee, Dipak Ghosal, "Wireless sensor network survey", Computer

- Networks, Vol. 52, No. 12, pp. 2292-2330, August 2008.
- [2] R. Szewczyk, E. Osterweil, J. Polastre, M. Hamilton, A. Mainwaring, D. Estrin, "Habitat monitoring with sensor networks", Communications of the ACM, Vol. 47, No. 6, pp. 34–40, 2004.
- [3] S. H. Chauhdary, A. K. Bashir, S. C. Shah, M. S. Park, "EOATR: energy efficient object tracking by auto adjusting transmission range in wireless sensor network", Journal of Applied Sciences, Vol. 9, No. 24, pp. 4247–4252, 2009.
- [4] P. K. Biswas, S. Phoha, "Self-organizing sensor networks for integrated target surveillance", IEEE Transactions on Computers, Vol. 55, No. 8, pp. 1033– 1047, 2006.
- [5] D. Kumar, "Monitoring forest cover changes using remote sensing and GIS: a global prospective", Research Journal of Environmental Sciences, Vol. 5, pp. 105–123, 2011.
- [6] Efficient Cluster Head Selection Methods for Wireless Sensor Networks, journal of networks, vol. 5, no. 8, august 2010
- [7] Energy Efficient Routing Algorithm with sleep scheduling in Wireless Sensor Network Subhash Dhar Dwivedi et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (3) , 2012, 4350 - 4353